DOCUMENT RESUME

ED 125 301 FL 007 837

AUTHOR Wilbur, Ronnie B.; Menn, Lise
TITLE Towards a Redefinition of Psychological Reality: On the Internal Structure of the Lexicon. San Jose Occasional Papers in Linguistics, Vol. 1.
INSTITUTION San Jose State Univ., Calif.
PUB DATE 75
NOTE 11p.; Paper presented at California Linguistics Association Conference (San Jose State University, Summer 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Cognitive Processes; Definitions; Dictionaries; Language Development; Language Patterns; *Linguistic Theory; Morphemes; *Morphology (Languages); *Psycholinguistics; Receptive Language; *Semantics; *Vocabulary; Vocabulary Development

ABSTRACT

Evidence for speaker knowledge of morphological patterns, both derivational and inflectional, is not limited to productive patterns. Nonproductive patterns appear to be accessible in such a way that accessibility (a term preferred to "psychological reality") may be viewed as a function of four somewhat interdependent factors: (1) productivity, (2) semantic transparency (e.g., "drunk-drunkard" is nonproductive but transparent), (3) morphological paradigmaticity (membership in a traditional inflectional paradigm), and (4) phonological relatedness. Each of these factors varies along a continuum. Highly accessible but nonproductive patterns may be overgeneralized, e.g., "bring, brang, brung." To incorporate information about pattern accessibility into a formal grammar, it is proposed that the speaker's knowledge of unproductive patterns be captured by the introduction of a sublexical level of morphemic analysis representing the maximal redundancy recognizable to speakers. The lexical level contains all information not predictable from productive rules; a lexical entry is related to a sublexical entry by sublexical "derivational" rules, but this decomposition and derivation is to be interpreted only as word analysis, not as word synthesis. This addition to morphological theory is illustrated with the nonproductively related Spanish words "puerta" and "porton." It is also noted that English speakers have some access to meanings encoded by nonproductive morphological process, both Latinate and English, when they are asked to guess "meaning" of properly made up nonwords in a multiple-choice task. (Author/DB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED125301

FL007837

Towards a Redefinition of Psychological Reality: On the Internal Structure of the Lexicon *

Ronnie B. Wilbur and Lise Menn

The goal of the linguist is to construct a grammar which characterizes the competence of the ideal speaker/hearer. It nonetheless remains a legitimate question to ask what the relationship is between the rules linguists write and the processing which takes place in the mind of the speaker. It may be argued that linguistic theory should attempt to produce a working model of language, but that the formalism employed therein need not necessarily reflect the knowledge of the adult speaker. This argument can be appropriately countered by pointing out that if several possible grammars which account for the data can be postulated, one which reflects what the speaker knows is preferable to any others. This point is particularly relevant to the question of whether the speaker actually derives forms in his head by means of derivations which contain morphological and phonological rules. Thus it has been one of the goals of psycholinguistics to search for evidence for derivational processing in speakers (MacKay, 1974). If this search produces no positive support for a derivational model, then presumably one would take recourse to a lexical model in which all the words are listed in the lexicon and related by passive generalizations, as for example proposed by Venneman (1974). Since psychological reality is the crucial factor in testing any model which we propose, it is necessary to step back and consider what psychological reality really is.

"Psychological reality" has been taken by many linguists as synonymous with "rule productivity". This has resulted in tests for productivity (does the rule apply to new or loan words? to nonsense words? does the rule predict systematic patterns in hypercorrections, overgeneralizations, language learning errors, etc.?) being considered also as tests of psychological reality. We have previously argued that psychological reality and productivity are separate concepts (Wilbur and Menn, 1974). In this previous discussion, we pointed out that generative rules serve a double function--first to generate in dynamic derivations several surface forms from a single underlying form (thus reflecting the relationship between groups of words) and second to capture generalizations about the relationship between words without actually deriving the words. This difference can be reflected in two types of notation. The first function of rules can be written as $A \rightarrow B / X __ Y$ while the second is more appropriately formalized as *XAY. The notation *XAY is intended to reflect the fact that the generalization is true of the language even though no productive rule exists to create the situation. Patterns which are non-productive may nonetheless be accessible to the speaker for use in analogy, new formations, overgeneralizations, etc. For example, essentially closed sets, such as the pronoun system of most languages, which can without question simply be learned as lists, show operation of analogy, (my:mine, your:yourn, his:hisn). This may be taken as evidence that patterns can be abstracted by speakers even in closed subsets of the lexicon. Thus there are patterns which are available to the speaker which can hardly be considered productive rules. Productivity is a feature of some rules of the language (language being used here in an idealized sense, in that is exists independent of its usage by particular speakers) while psychological reality is the extent to which language patterns are reflected in the usage of that language's speakers.

---

* appeared in San Jose Occasional Papers in Linguistics, 1975, Vol. 1, pp. 212-221.

We prefer to refer to the accessibility, availability, or utility of a pattern to the speaker rather than to use the term psychologically real, which already has several possible (mis)interpretations to it. How is this accessibility to be characterized? Initially, we hypothesized that it consisted of two gradients, one which covers a range of productivity, and the other which covers a range of semantic transparency (See Appendix). The productivity gradient is a continuum between one pole of full productivity and the opposite pole of complete lack of productivity. This continuum is a direct reflection of the degree to which a particular pattern contributes to the redundancy of the lexicon (the more productive, the more redundant). The semantic transparency gradient ranges from opaque (canine-hound, century-hundred, etc.) through translucent (weal-wealth, steal-stealth, heal-health) to transparent (drunk-drunkard, talk-talked, retain-retainer), and refers to the extent to which a semantic correspondence is still synchronically visible. Presumably, dead patterns are available only to those with formal knowledge of former stages of the language or of related languages. And presumably productive patterns are available for analogizing or overgeneralizing without conscious awareness on the part of the speaker.

In addition to these two gradients, two other factors are involved in pattern accessibility. One of these is obvious--phonological awareness, which is itself a continuum. Certain patterns will be more visible simply because the phonological system, while other patterns may be more or less visible (dear-darling, wring-wrench, stink-stench, cling-clench, etc., retain-retention, sing-sang-sung). The other factor related to pattern accessibility is morphological paradigmaticity, the force that binds talk and talked, sing and sang, catch and caught, go and went as present and past tenses of the same verb, for example. In general, morphological paradigmaticity may be to inflectional relations what semantic transparency is to derivational relations. For the present, we will retain both factors, although it may be the case that they represent two ends of an inflectional/derivational continuum, as there are analyses of languages in which it is impossible to state definitively whether you are inflecting or deriving.

What we are claiming then is that pattern accessibility ("psychological reality") is a 4-factor function: 1) productivity, 2) semantic transparency, 3) morphological paradigmaticity (Saltarelli and Calvano, 1974), and 4) phonological relatedness. Each of these factors is itself a continuum and they are of course interdependent. However, no one factor by itself is sufficient. A highly productive rule must by its nature include at least one of the other factors, that is, it must be a highly productive semantic rule, or a highly productive morphological rule, or a highly productive phonological rule. It appears to be the case that less productive or marginally productive rules require middle to high ratings at at least two of the remaining factors. (If semantic transparency and morphological paradigmaticity are viewed as being on a single continuum themselves, then we have only three factors and we can say that for a pattern to be accessible, middle to high ratings are needed on two of the three factors.)

What is the potential benefit of such a definition? It allows us to account for degrees of availability of patterns, while taking into account degrees of productivity, degrees of semantic relatedness, degrees of phonological relatedness. It allows us to account for why some mistakes are more likely than others. Consider the following example.

A frequent overgeneralization of children learning English is to conjugate bring as bring, brang, brung or as bring, bringed, bringed. The addition of the -ed suffix on bringed causes no difficulty for current theory since we feel comfortable in calling the addition of -ed a productive rule in English. However, the occurrence of bring, brang, brung is taken by some linguists as evidence for its "psychological reality." Do we really want to say that English has a productive rule of vowel ablaut which is used to form the past tense and past participle of some verbs? Is it really productive if it overapplies to 1 or 2 verbs (including think)? It is not really reasonable to assign English a synchronic rule of vowel ablaut on the basis of 1 or 2 overgeneralizations. Yet there is a sense in which the ablaut overgeneralization is a more likely mistake than for a verb to be assigned to the fight, fought, fought class. Let us consider how this difference in probability of occurrence is reflected in the definition which we have proposed. We would submit that both patterns are synchronically unproductive. Because of their participation in a morphological paradigm, both patterns are equivalent semantically. The morphological paradigm is the same for both, that is, present-past-past participle. In terms of phonological relatedness, however, the ablaut alternation is higher up on the scale than the -ought, -ought alternation for two reasons. One is that there is a greater degree of phonological similarity between the present, past, and past participle in the ablaut forms since only the vowel is changed, than there is in the -ought forms where generally only the first consonant is retained (catch-caught, teach-taught, buy-bought, etc.). The other is that there is a greater degree of phonological coherency among the class of verbs which are subject to ablaut than among the -ought verbs. That is, one can extract a phonological generalization which binds the ablaut verbs and allows one to assign a new verb to this group, namely that verbs of the shape /(s)Cri(g,k)/(drink, stink, sink, sing, ring, etc.). With the -ought verbs, there is no phonological generalization which one can use to assign a verb to this class (teach, catch, buy, bring, fight, seek, think).

Within the current framework, the occurrence of brang is taken as evidence that ablaut has extended its domain to other verbs. Here, its existence may be viewed as evidence for a higher saliency of the pattern due to the greater phonological coherency of the group of verbs which undergo the rule, so that assignment to the ablaut class can be made on a synchronically available generalization, whereas assignment to the -ought class seems not to have any synchronic motivation. Thus we can speak of both patterns as being unproductive in the language, and also predict that the ablaut pattern will be more likely to be used by speakers for analogy than the -ought pattern. This avoids the problem of saying a rule doesn't exist because only a small percentage of speakers use it and the large majority don't. It allows us to characterize the behavior of the small percentage (the fact that they all behave similarly with respect to a particular set of words, that they had to get this behavior from somewhere, that their behavior mirrors a formerly productive stage of the language) without claiming synchronic productivity for the rule.

In the standard theory, we are forced to strive for a single underlying representation which can generate as many related words as possible, utilizing different rules in each derivation. Deriving them from a single source incorrectly implies a productive rule. The alternative, listing them in the lexicon, leaves many important relationships uncaptured. Vennemen's (1974)

approach of listing everything and relating groups by via rules fails to capture differing degrees of likelihood of usage. We would like to suggest an appropriate compromise.

It is clear to us that the use of a morphological or phonological rule to actually derive a set of words from a single common source is appropriate for processes which can be considered as truly synchronically productive. The problem to be resolved is what the appropriate way is to handle groups of words which may be related by processes which are lower on the productivity scale, but high on one or more of the other scales and thus still reflect a valid synchronic relationship but which for semantic, morphological, or phonological reasons result in "derivations" which we believe ought to be considered undesirable. We would like to propose that a word be permitted to have two source forms which are relevant to it, sublexical and lexical.

The sublexical representation represents the maximal decomposition of a word which can be synchronically justified. The lexical representation contains all of the non-predictable information (and probably some of the predictable information as well). The two levels are related by non-productive sublexical rules which do not synthesize, but rather provide an analysis of the word. These sublexical rules are the ones that are nice to know, and presumably only the ideal speaker/hearer knows them all. The added advantage of the sublexical level is that it allows us to talk about the psychological reality (saliency, accessibility, utility, availability) of patterns which we readily acknowledge are unproductive.

Consider "exponent", a word containing three Latinate morphemes. The ideal speaker/hearer knows that "exponent" is polymorphemic; the naive high school algebra student does not. We can capture the fact that this information is available to the ideal speaker/hearer by giving "exponent" the sublexical entry ex+pon+ent(agent). Yet we wish to indicate that "exponent" is not on a par with other polymorphemic words derived by productive processes of English, like quickly or quitter. The + junctures in ex+pon+ent (agent) do not represent "live" junctures and may well be removed by sublexical rules which relate the sublexical forms to the lexical level. We can capture the fact that a certain sense of the identity of -ent as an agentive noun or adjective suffix is available to speakers by leaving the sublexical + boundary preceding "ent" in at the lexical level by labeling it as an agentive suffix only at the sublexical level. It will differ from fully productive agent formations in -er because these will not even have to be listed in the lexicon, and from non-productive agentive formations like butcher because these will have a lexical representation in which -er is joined with a + boundary and also labeled with its meaning. In other words, a fully transparent morpheme that must be listed in the lexicon is listed there with its meaning; a less transparent morpheme will have its meaning listed only sublexically; and morphemes for which one needs training in historical linguistics in order to relate on a meaningful level are not assigned any common meaning at either the sublexical or the lexical level, as they only have meaning insofar as they participate in whole words in the language but none by themselves. This type of formulation provides a means of accomodating a range of semantic transparency.

That the meaning of certain (but not all) morphemes which no longer participate in productive derivations is available to speakers and should thus be listed somewhere in the lexicon (even though they are very close to opaque) can be seen by the following results which indicate that speakers can agree on the possible meanings of words made up of these morphemes even though they probably couldn't (if asked) predict them. Native speakers were asked to choose the most likely meaning (of three) for eleven different words created with various stems and affixes which are no longer transparent enough to be given meanings at the lexical level, nor productive enough to have their own derivational morphological rules. Table I summarizes the results by indicating the percent of people who chose each of the three responses. Subjects were all undergraduate students enrolled in a first-year linguistics course.

TABLE I Percent agreement, Native speakers N=73

Stimulus	Meaning a	%	Meaning b	%	Meaning c	%	
abducive	informing	18.38	distracting	71.84	conserving	9.86	*
supponent	fundamental	59.72	excessive	27.78	softening	13.89	*
erigible	capable of being: twisted	38.03	raised	29.58	bent	30.99	NS
egredient	that which is: smoothed	34.78	held	39.10	vaporized	26.09	NS
arrective	raising	47.14	retarding	22.86	impairing	30.00	*
degressive	going down	83.56	coming up	4.11	turning around	12.32	*
desponsive	chastening	11.27	cheering	1.41	saddening	87.32	*
obsist	blockade	75.00	melancholy	13.24	relate	11.76	*
sustension	holding steady	91.78	running quickly	4.11	walking noisily	4.11	*
retent	spent	7.04	kept	85.92	spoken of	7.04	*
chibble	light rain	22.22	a kind of smooth cloth	12.36	coarse sawdust	65.28	*

* Significant at p .01

The results indicate that enough of the meaning of the various stems and affixes (with two exceptions) is available to speakers that some indication of their meaning independent of the words they occur in should be indicated in the grammar. We suggest that the sublexical level is the appropriate place for this to be done.

The following problem from Spanish illustrates our conception of a sublexical "derivation" and its relation to the lexical level and to productive phonological rules. The forms puerta "door," puertita "little door," and porton "gate" are clearly related, but attempts to describe this relationship in a synchronically dynamic manner leads to a rule ordering paradox between stress shift and diphthongization. In porton, stress must be shifted off port and onto on before diphthongization applies, since diphthongization causes stressed vowels to diphthongize and would incorrectly produce *puerton. In puertita, the stress must not be shifted off port and onto it until after diphthongization has created puert. The diminuitive formation is still productive in Spanish, but the augmentative formation is totally dead. Porton may not be seen by some speakers as related to puerta and is therefore semantically translucent. The problem is how to account for the non-application of the diphthongization rule in the form porton.

Sublexical representation[2]	/port + /	----	/port + /
Augmetative formation	----	----	port + on
Boundary adjustments	----	----	port on
Lexical representation	/port + a/	----	/porton/
	=	nom fem 'door'	nom masc 'gate'
Diminuitive formation	----	port + it + a	----
Diphthongization	puert + a	puert + it + a	----
Stress Shift	----	puert + it + a	----
Output	puerta 'door'	puertita 'little door'	porton 'gate'

To obtain the correct surface forms, stress must be shifted off the stem of porton before diphthongization applies. The rule which shifts the stress appears to be the same rule that shifts the stress in puertita. In fact, what we are seeing is a reflection of a previous historical stage in which the augmentative-creating rule was productive and the stress shift rule applied productively to its output. The augmentative rule is no longer productive, and the stress shift rule does not "apply" productively to its output. We are claiming that the speaker knows porton as a whole word with the stress on the second syllable. The speaker may analyze porton into port + on and can find a relationship between it and puerta. But he doesn't have to in order to produce porton correctly while speaking. In fact, he could just as well spend his entire life without ever realizing that porton and puerta are related, and still use each word appropriately. On the other hand, to know what the diminuitive means (and to know its gender and number), he has to know the source form.

Consider an alternative analysis. Harris (1969) uses a phonological cycle to formally account for these forms. When the words enter the phonological component, they are bracketed as: [[port +a] [port + on] but [[[port] + it] + a]. On the first cycle, stress shift applies within the first set of brackets, shifting the stress to the -on suffix in porton. Diphthongization then applies to port + a and port + it+ a to give puert + a and puert + it + a. On the second cycle, stress shift applies again, shifting the stress to the suffix in puert + it + a. There is no trouble here with the mechanics. The problem is motivating the placement of the brackets. They give the right output, but they make a number of incorrect claims. 1) They imply that -on is on a par with be dimunitive suffix, which is certainly incorrect. In other words, this analysis does not capture differences in the productivity of the augmentative and diminutive formation rules. 2) They imply that the meaning of porton should be as predictable as puertita is from puerta. This analysis does not capture differences in meaning. 3) They imply that the speaker is as likely to err in the direction of *portita as in the opposite direction of *puerton. This is clearly an empirical claim but we do not know of any data relevant to it. Our analysis claims that the error puerton, if made at all, would be more likely than portita, since puertita is more closely related morphologically to puerta and is therefore more likely to resist reanalysis along the lines of porton.

It is possible to assign the placement of the brackets for such an analysis by adopting the convention that morphemes which appear at the lexical level with a + (or stronger) boundary are to be enclosed in brackets. So porton has no internal boundary and therefore no internal bracketing. Puertita is formed postlexically from the bracketed puert-. This is an acceptable solution within the standard framework, but lacks explanation for the placement of the boundaries in the first place. It is here that the addition of a sublexical representation becomes important. We want to say on the one hand that since the augmentative formation is totally non-productive, porton is treated by the speaker as a whole word, that the rules of Spanish treat it as a whole word, that synchronically in Spanish it exists with the stress on the second syllable, and that consequently there is no justification for positing a synchronic morpheme boundary on it. On the other hand, we want to say that a speaker can analyze porton into two parts, that it is not an accident that these two parts are port and on and not some other division such as porto and -n, that the semantic relationship between puerta and porton is available to a speaker and that -on still carries some predictive power with respect to possible meanings of a word which ends in it, and that therefore the speaker has some access to information about the internal structure of porton. The sublexical entry includes the information that porton can be divided into port + on, i.e. where the boundary is, and also that the port that occurs in porton is the same port that puerta comes from and that -on is the same -on that was added to form augmentatives. The sublexical entry includes information which the speaker has access to but does not necessarily need in order to be a native speaker of the language. The sublexical level includes redundancies about the structure of the language which are available to the speakers but which are not necessarily used by the speakers. Presumably speakers differ on the extent to which they make use of this sublexical information. A speaker of Spanish may never connect porton and puerta. He may only do so when it is pointed out. Yet another speaker may actively search for connections between words and may use them to organize (or reorganize) the internal structure of this lexicon. We are claiming

then that the sublexical level includes information about the structure of the language being learned or used, that this information is potentially available to the speaker, but that not all the information available is actually used, that speakers differ in the amount of information that they use from the sublexical level, and that this difference is reflected in individual lexical entries. The differences in the lexical entries will have to be reasonably constrained in order to reflect the fact that a "language" does exist in some coherent form somewhat independent of any particular speaker's grammar, just as the notion "circle" exists independent of any representation of it so long as the representation stays within some indefinable bounds of "circleness". We believe that when the concepts of lexical and sublexical entries are more carefully formalized, including the necessary parameters and their possible ranges, the framework of variable accessibility and sublexical structure will be able to account for: 1) the speaker's ability to constrain the possible meanings of words based on a Spracingefuhl of the meaning of its component morphemes, 2) the speaker's ability to analyze words into subparts that reflect earlier stages of the grammar, 3) speaker's ability to come up with the wrong analysis (folk etymologies) and restructuring (if all the information were clearly available to them, they wouldn't make mistakes), 4) range of speaker behavior (some know it, some don't) by separating the description of what's available to the speaker from what the speaker actually has, 5) "education" and the effects of "literary reading" (adding more sublexical rules, but not productive rules, to the grammar), and 6) different strengths of different boundaries (or put another way, the different visibility of different boundaries) and their effects on the application of phonological rules. The framework makes a number of empirical predictions which can be tested out on more and more words with more and more speakers.

Footnotes

1. The questionnaire "It Pays to Increase Your Word Power" from which Table 1 derives was developed for Wilbur and Menn (1974) as a demonstration that partial meanings of certain morphemes were still available to speakers in the context of being able to constrain the possible meanings of words containing those morphemes. It is not intended as a systematic exploration of the componenet morphemes, i.e. whether it is the de or the gress that carries the meaning, or whether one of the parts has a partial meaning and the other has none, or any other possible questions which can be raised in connection with the method. We recognize that certain of the words are more transparent in their relationships than others, retent (retention), sustension (sustain), but egredient(ingredient), chibble (the dimunuitive of chip). We submit that 1) the word game provides a method for investigating partial meanings, and 2) full-fledged investigations must control for the different kinds of relationships which can be explored, different degrees of relatedness, transparency, etc.

2. We are not making any claims as to the exact status of the nominative ending -a here. It could be that the lexical entry is /port+/ and a rule of nominative function adds the feminine ending -a.

We are also not making the claim that stress is phonemic on /port+/ at the sublexical level, only that it bears no abnormal relationship to the rules of Spanish stress placement, whereas porton does, in that it is frozen with the stress on the final syllable.

References

Harris, J. (1969) Spanish Phonology. Cambridge, Mass.: MIT Press.

MacKay, D. (1974) "On the Internal Structure of the Lexicon." Paper presented at the Winter Meeting, Linguistic Society of America, New York.

McCawley, J. (1975) Acquisition Models as Models of Acquisition.

Saltarelli, M. and Calvano, W. (1974) "The Morphological Justification of Phonological Rules." Presented at the Winter Meeting, Linguistic Society of America, New York.

Venneman, T. (1974) "Words and Syllables in Natural Generative Phonology." In Burck, A. et al.(eds.) Papers from the Parasession on Natural Phonology. Chicago: Chicago Linguistic Society.

Wilbur, R.B. and Menn, L. (1974) "The Roles of Rules in Generative Phonology." Paper presented at the Summer Meeting, Linguistic Society of America, Amherst.

PRODUCTIVITY GRADIENT

Productive----------Non-productive but expressing salient generalization----------obscure but expressing minor generalization----------Dead

CAPABLE OF CONTRIBUTING TO THE REDUNDANCY OF LANGUAGE

For all speakers----------For most speakers----------For some speakers----------For no speakers

TRANSPARENCY GRADIENT

Inflection----------Derivation

Fully transparent
Only irregular forms found in dictionaries

Productive----------Less Productive Transparent.....
May also have idiomatic meaning

Productive----------Unproductive
Isolated examples may be transparent, e.g. "drunkard"

.....May be transparent in restricted semantic areas, e.g. "heighth" (non-std.).....

.....Translucent.....
Analogous new forms would be understood but considered odd. (Some found in advertising.)

Not generally available for creative conscious analogizing. Sound plausible when analyzed. Still provide redundancy, e.g. "lovely", "steal-stealth"

.....Opaque.....
Entered in dictionaries. (Presumably) dead correspondences like "cardiac-heart", "canine-hound", "century-hundred", "cannabis-hemp"

APPENDIX (Adapted from Wilbur and Menn, 1974)